

Université Montpellier II
Ecole Doctorale Biologie Integrative

Use Body
non numérisé

Diplôme d'Etudes Approfondies de Parasitologie

2002-2003

Diversité nucléotidique chez des populations naturelles d'*Anopheles gambiae* s.s., vecteur de paludisme en Afrique.

Par Nicolas PONÇON

Laboratoire d'accueil

Laboratoire de Lutte contre les Insectes Nuisibles (LIN)

UR 016 "Caractérisation et contrôle des populations de vecteurs"

Institut de Recherche pour le Développement (IRD)

Montpellier

22 DEC. 2003

Responsable de stage

Didier FONTENILLE



052
ANOPAL
PON

MF_A



SOMMAIRE

1. INTRODUCTION	1
1.1. GENERALITES	1
1.1.1. Le parasite	1
1.1.2. Le vecteur	1
1.1.3. Le cycle de Plasmodium	2
1.1.4. La lutte	2
1.2. ETAT DES CONNAISSANCE SUR LA COMPETENCE VECTORIELLE	3
1.2.1. Bases génétiques de la transmission	3
1.2.2. Les données du génome	3
1.3. OBJECTIF DE CETTE ETUDE	4
2. MATERIEL ET METHODES	5
2.1. LES GENES ETUDIES	5
2.1.1. Choix des gènes	5
2.1.2. Recherche des séquences nucléotidiques	6
2.1.3. Détermination des amorces PCR	6
2.2. MATERIEL BIOLOGIQUE	7
2.3. METHODE	7
2.3.1. Extraction de l'ADN	7
2.3.2. Amplification des séquences des gènes par PCR	8
2.3.3. Electrophorèse sur gel d'agarose	9
2.3.4. Séquençage	9
2.3.5. Analyse des données	9
3. RESULTATS	10
3.1. AMPLIFICATION ET SEQUENCAGE DES GENES CIBLES	10
3.1.1. Amplification des séquences	11
3.1.2. Séquençage	11
3.2. BIAIS DANS L'USAGE DES CODONS	12
3.3. POLYMORPHISME NUCLEOTIDIQUE	12
4. DISCUSSION	13
4.1. BIAIS DANS L'USAGE DES CODONS	13

4.2. POLYMPORPHISME NUCLEOTIDIQUE	14
5. CONCLUSION	17
TABLEAU 1, TABLEAU 6	18
TABLEAU 4, TABLEAU 5	19
TABLEAU 7	20
PROJET DE RECHERCHE	21
1. INTITULÉ	21
2. CONTEXTE	21
2.1. Présentation de la maladie	21
2.2. Importance	21
2.3. Epidémiologie	22
2.4. Les vecteurs	22
2.5. Aedes vexans	22
3. OBJECTIF	23
4. PROTOCOLE	23
4.1. Zones étudiées	23
4.2. Etude de certains caractères biologiques	23
4.3. Etude de la compétence vectorielle	24
4.4. Etude de la systématique et de la structuration des populations d' <i>Ae. vexans</i>	24
5. RESULTATS ATTENDUS	25
6. APPLICATIONS	25
BIBLIOGRAPHIE	26

Diversité nucléotique chez des populations naturelles d'*Anopheles gambiae* s.s., vecteur de paludisme en Afrique.

Anopheles gambiae s.s. est le vecteur majeur du paludisme en Afrique. L'efficacité de la transmission de *Plasmodium* dépend de la compétence et de la capacité vectorielle de ce moustique. Il a été démontré que la compétence vectorielle d'*An.gambiae* et certains facteurs de la capacité (comme l'anthropophilie) avaient un déterminisme génétique. Dès lors, l'identification de polymorphismes génétiques associés à certains phénotypes tels que la permissivité vis à vis du *Plasmodium* ou l'anthropophilie est devenu un nouvel enjeu.

L'objectif de ce travail visait à étudier le polymorphisme nucléotidique d'*An.gambiae* à partir de 3 populations naturelles, les moustiques provenaient du Sénégal, du Burkina Faso et du Cameroun. Le polymorphisme a été recherché au niveau de 16 gènes. Ces gènes appartiennent à des catégories fonctionnelles différentes: gènes impliqués dans la réponse immunitaire (reconnaissance et modulation du signal), récepteurs constitutifs de l'intestin, gènes associés à l'olfaction et gènes de maintenance liés au métabolisme général. Des moustiques des 2 formes moléculaires M et S ont été criblés. L'analyse globale a porté sur 8 de ces gènes. Nos résultats ont montré un biais dans l'usage des codons, le biais est dirigé vers une préférence des codons se terminant par G et C et il est plus important pour les gènes de maintenance. Nous avons observé une diversité nucléotidique moyenne de 17 SNPs (Single Nucleotide Polymorphism) par kilobase. La diversité est plus élevée pour les gènes Sp14D1, Sp14A, TEP3 qui sont impliqués dans l'immunité et pour laminin, récepteur constitutif de l'intestin des anophèles auquel se lie le parasite. Ces 4 gènes pourraient être soumis à une sélection adaptative en zones palustres, et constituent donc de bons candidats pour des études d'associations génotype-phénotype. Enfin, pour le gène TEP3, une mutation apparaît fixée entre les formes M et S du Burkina Faso, ce qui pourrait sous-entendre une adaptation spécifique localité-dépendante.

Mots-clés : *Anopheles gambiae* s.s., vecteur de paludisme, diversité nucléotidique, immunité, sélection adaptative.

Nucleotide diversity of wild populations of *Anopheles gambiae* s.s., malaria vector in Africa.

Anopheles gambiae s.s. is the most efficient and the most common malaria vector in Africa. *Plasmodium* transmission depends on vectorial competence and capacity. It has been shown that competence and some capacity components (such as host preferences) are genetically determined. Thus, identifying genetic polymorphisms associated with specific phenotypes (permissiveness for exemple) appears very interesting.

We studied nucleotide polymorphism of *An.gambiae* among 3 different wild populations. Mosquitoes used in this study were of molecular forms M and S, they were collected in Senegal, Burkina-Faso and Cameroon. Polymorphism was investigated for 16 genes belonging to different functional groups : genes associated to immunity (recognition or signal modulation), constitutive gut receptors, odorant receptors and housekeeping genes. Complete analysis has concerned 8 genes. Our results showed that codon usage bias favors G and C ending codons, and that highly expressed genes were the most biased genes. The average nucleotide diversity was 17 SNPs (Single Nucleotide Polymorphism) per kilobase. Sp14D1, Sp14A, TEP3 (genes with a putative role in parasite defense) and laminin (specific gut receptor for *Plasmodium*) exhibited significantly higher levels of polymorphism. This suggests that these 4 genes could be under selective constraints. Furthermore, TEP 3 showed great evidence for a fixed SNP between the M and S forms in the Burkina Faso, that may result of a location-dependant specific adaptation.

Key-words : *Anopheles gambiae* s.s., malaria vector, nucleotide diversity, immunity, positive selection.

1. INTRODUCTION

1.1. GENERALITES

Malgré les efforts de lutte entrepris depuis plus de 50 ans, le paludisme demeure la plus grande endémie parasitaire mondiale. L'Organisation Mondiale de la Santé estime l'incidence mondiale de la maladie à 300-500 millions de cas cliniques par an et évalue le nombre de victimes entre 1 et 3 millions chaque année. Plus de 90% des cas de paludisme surviennent en Afrique subsaharienne et les décès suite à l'infection sont essentiellement comptés parmi les enfants de moins de 5 ans [1].

1.1.1. Le parasite

Le paludisme est une maladie à transmission vectorielle causée par un parasite hématozoaire du genre *Plasmodium*, appartenant au groupe des Apicomplexa. Le genre *Plasmodium* regroupe plus d'une centaine d'espèces, toutes parasites, dont quatre sont spécifiques de l'homme : *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale*. *P. falciparum* est l'espèce la plus répandue dans tous les pays intertropicaux et est responsable de 99% des infections palustres mortelles [2].

1.1.2. Le vecteur

Tous les *Plasmodium* de primates, dont les 4 espèces parasites de l'homme sont transmis par des moustiques du genre *Anopheles* (famille des Culicidae, ordre des Diptères). Le genre *Anopheles* comprend environ 400 espèces dont une soixantaine seulement est vectrice de *Plasmodium*. Vingt espèces sont à l'origine de la plupart des cas humains [3,4].

La région afro-tropicale (Afrique subsaharienne et Madagascar) est une zone endémique, où la transmission palustre est assurée par cinq espèces majeures. Les espèces anophéliennes appartiennent à des complexes d'espèces, chaque complexe regroupant différentes espèces indifférenciables morphologiquement.

- *Anopheles gambiae* et *Anopheles arabiensis*, fréquemment présents en sympatrie, sont deux vecteurs très répandus et très efficaces. Ces espèces appartiennent au complexe *Anopheles gambiae* qui regroupe 7 espèces cryptiques reproductivement isolées. Au sein de l'espèce *An. gambiae s.s.*, on distingue deux formes moléculaires différentes : M et S. La caractérisation moléculaire des 2 formes s'effectue à partir du polymorphisme de l'ADN ribosomal. Actuellement, les croisements fertiles entre formes moléculaires M et S sont possibles en conditions de laboratoire, mais les hybrides sont extrêmement rares,

voire absents, en zone de sympatrie, ce qui laisse supposer que M et S sont deux espèces en voie de spéciation.

- *Anopheles funestus* est la seule espèce du complexe *Anopheles funestus* impliquée dans la transmission palustre humaine [5].
- *Anopheles nili* et *Anopheles moucheti* sont deux espèces qui ont une importance régionale, notamment dans certaines régions densément peuplées des savanes humides et des zones équatoriales [6].

1.1.3. Le cycle de Plasmodium

Le cycle des *Plasmodium* humains se déroule successivement chez l'homme, hôte intermédiaire, et chez l'anophèle femelle hématophage, hôte définitif. La femelle, au cours d'une piqûre infectante, injecte des sporozoïtes, forme infectante de la plasmodie pour les vertébrés, à l'homme. Ces sporozoïtes entament ensuite, après un cycle pré-érythrocytaire (et formation d'hypnozoïtes pour *P.vivax* et *P.ovale*) dans le foie, une série de cycles schizogoniques intra-érythrocytaires. Chaque schizonte libère un grand nombre de mérozoïtes qui peuvent réinfecter des hématies. Après un nombre variable de générations schizogoniques, des mérozoïtes évoluent en gamétocytes mâles et femelles (la forme sexuée du parasite) qui apparaissent dans le sang périphérique de l'homme. La femelle moustique s'infecte au cours d'un repas de sang sur un porteur de gamétocytes. Au cours du cycle sporogonique, les gamétocytes évoluent en macrogamètes femelles et microgamètes mâles qui, suite à la fécondation, donneront un zygote. Ce dernier se transforme en ookinète qui traverse la membrane péritrophique puis l'épithélium intestinal pour s'enkyster sous forme d'oocyste au niveau de la membrane basale de l'intestin du moustique. Cet oocyste, après avoir subi des divisions sporogoniques, libère des sporozoïtes dans l'hémocèle du moustique. Ceux-ci gagnent alors les glandes salivaires de la femelle, qui devient ainsi infectante. Le cycle sporogonique dure de 7 à 10 jours selon l'espèce de *Plasmodium* et la température [4]. Au cours des différentes phases de ce cycle, on observe une réduction parasitaire drastique. L'intestin et les glandes salivaires représentent les compartiments où les réductions sont les plus importantes [49].

1.1.4. La lutte.

La lutte actuelle contre le paludisme s'oriente selon deux axes :

- la lutte contre le parasite : elle repose sur l'utilisation de la thérapie combinée et vise à favoriser l'accès aux structures de soins

- la lutte contre le vecteur : elle préconise l'utilisation de moustiquaires imprégnées et les aspersions intra-domiciliaires d'insecticides. Chacune de ces actions, pour réduire efficacement la morbidité palustre, doit être adaptée au niveau de transmission, à la prévalence, aux espèces vectrices et aux conditions épidémiologiques propre à chaque foyer [7].

L'émergence de résistances aux antipaludéens chez *Plasmodium* ainsi qu'aux insecticides chez le moustique a fortement entravé les opérations de lutte et de nouvelles stratégies de lutte basées sur une meilleure connaissance des relations *Plasmodium*-vecteur doivent être développées.

1.2. ETAT DES CONNAISSANCE SUR LA COMPETENCE VECTORIELLE

1.2.1. Bases génétiques de la transmission

Pour être considéré comme vecteur efficace de plasmodies humaines, un anophèle doit présenter une capacité vectorielle suffisante. Celle-ci exprime à la fois le degré de compatibilité entre le vecteur et le parasite (compétence vectorielle) et le fonctionnement de ce système dans un environnement donné. Elle dépend donc de la réceptivité du vecteur au parasite, de ses préférences trophiques, de la fréquence des repas, de la longévité du vecteur...[4]

La sélection d'une souche d'*An. gambiae* réfractaire à l'infection a permis de démontrer les bases génétiques de la transmission chez le vecteur. Le phénotype réfractaire de cette souche se caractérise par l'encapsulation des oocystes sur la membrane basale de l'intestin [42]. Trois locus de trait quantitatif (QTLs) ont été identifiés comme associés à la capacité d'encapsulation [43]. Il a également été démontré que le moustique développe une réaction immunitaire au cours de l'infection. En effet, des peptides immuns (GNBP, NOS, ICHIT) se sont révélés sur-exprimés au moment où les ookinètes traversent la barrière intestinale [45, 46]. Actuellement, avec le développement des puces ADN et des techniques de transgénèse, un nombre grandissant de molécules associées au développement du parasite chez le moustique est identifié [9,48,37]. Cependant, aucun gène bloquant le cycle sporogonique n'a été caractérisé.

1.2.2. Les données du génome

Le séquençage du génome d'*An. gambiae* a été achevé en octobre 2002 [41]. Il donne dorénavant accès à l'intégralité des séquences du génome du moustique à travers des banques

de données internationales. La séquence du génome a été obtenue à partir de la souche PEST d'*An. gambiae*. Cette souche est issue du croisement de colonies de formes moléculaires M et S et l'assemblage du génome a été rendu difficile du fait de la présence d'haplotypes doubles. Le génome est alors apparu très polymorphe, en effet un total de 444 963 Single Nucleotide Polymorphisms (SNPs) a été identifié [41]. Ces SNPs représentent les variations d'ADN les plus fréquentes dans le génome, leur nombre varie de 1 substitution pour 1000 paires de bases (pb) dans le génome humain à 10 substitutions pour 1000 pb chez la drosophile. Des travaux ont révélé que certaines maladies génétiques étaient liées à des mutations ponctuelles, par exemple la maladie de Crohn chez l'homme ou la tremblante chez les ovins, et les SNPs ont alors suscité un intérêt grandissant en génétique [47,35]. Ainsi, l'identification de polymorphismes nucléotidiques prédisposant à une pathologie devient un nouvel enjeu. Dans le cas d'*An. gambiae*, l'étude des polymorphismes nucléotidiques au niveau des gènes susceptibles d'être impliqués dans la compétence ou la capacité vectorielle, gouvernant des phénotypes comme la permissivité vis à vis du *Plasmodium* ou l'anthropophilie pourrait permettre d'identifier des gènes candidats à utiliser pour élaborer de nouvelles stratégies de lutte.

1.3. OBJECTIF DE CETTE ETUDE.

Ce travail a consisté à étudier la diversité nucléotidique d'*An. gambiae s.s* chez des moustiques issus de populations naturelles d'Afrique.

Le polymorphisme nucléotidique a été étudié, par comparaison de séquences, dans les régions codantes et non codantes des gènes ciblés. Nos analyses ont permis d'étudier l'usage des codons. Les taux de substitutions synonymes et de remplacements ont été analysés au regard de la fonction des gènes. Les résultats ont été rapportés à l'origine des populations échantillonnées afin d'identifier des substitutions spécifiques selon la localité d'origine ou la forme moléculaire des moustiques.

Le but était d'identifier des gènes dont le polymorphisme pourrait refléter une sélection adaptative. Ceux-ci constitueraient des gènes candidats judicieux pour mener des études d'association entre le polymorphisme nucléotidique et un phénotype donné tel que la permissivité d'*An. gambiae* aux *Plasmodium* ou l'anthropophilie.

2. MATERIEL ET METHODES

2.1. LES GENES ETUDIES

2.1.1. Choix des gènes

Nous avons choisi d'examiner des gènes impliqués dans le métabolisme général, des gènes potentiellement impliqués dans des mécanismes adaptatifs tels que les préférences trophiques ou les interactions vecteur/parasite et des gènes associés à la réponse immunitaire du moustique. Nous avons retenu :

- quatre gènes associés au métabolisme général: **RpL31** et **RpS11** qui codent pour les protéines ribosomiales, **G6PD** qui code pour la Glucose 6 Phospho Déhydrogénase et **NaK** codant pour un canal sodium/potassium.
- deux gènes impliqués dans les mécanismes de l'olfaction. Le gène **OR1** code pour un récepteur candidat, auquel se lient les molécules odorantes. Il est susceptible de jouer un rôle dans les préférences trophiques : son expression n'est induite que dans les antennes des femelles, et il est très réprimé 12 heures après un repas sanguin, période qui, chez le moustique, correspond à une diminution du comportement de recherche d'un hôte[11]. Le gène **arrestin 1** code pour un peptide régulant la transmission de signaux impliqués dans l'olfaction. Son action est assez large car il agit également au niveau de la transmission des messages visuels[12]. Il pourrait jouer un rôle dans la recherche de l'hôte nourricier ou du partenaire sexuel.
- deux gènes codant pour des constituants importants de l'intestin des anophèles. Vlachou et al. ont montré que **laminin γ 1**, très abondante au niveau de la membrane basale pourrait agir comme récepteur pour le ligand P25 des ookinètes de *Plasmodium berghei* chez *Anopheles gambiae*, permettant ainsi le développement de l'oocyste au niveau de la membrane basale [8]. La mucine **AgMuc1** est une protéine de surface très abondante sur les microvillosités intestinales [40]. AgMuc1 pourrait servir de récepteur pour *Plasmodium* au niveau du pôle apical des cellules intestinales. En effet, des analogues structuraux d'AgMuc1 sont impliqués dans la même fonction vis à vis de *Trypanosoma cruzi* chez les mammifères [40].
- quatre gènes de l'immunité, codant pour des peptides appartenant à la famille des PRR ("pattern recognition receptors") impliqués dans la reconnaissance de motifs "pathogènes". Les PRRs sont des récepteurs solubles ou liés à des cellules, qui peuvent faciliter la phagocytose (rôle d'opsonine), initier la transduction d'un signal ou déclencher

une cascade protéolytique. Christophides et al. ont montré que, suite à un challenge à *Plasmodium* chez *An. gambiae*, **TEP 4** et **PGRP-LB** sont induits rapidement et durant tout le cycle sporogonique, ce qui suggère une réponse systémique. A l'inverse, **TEP3** et **GNBP A1** ne le sont que tardivement, laissant penser qu'ils interviennent dans une réaction contre des stades plus tardifs du *Plasmodium* (oocystes, sporozoïtes) [9]. Ces 4 gènes de la famille des PRRs ont été retenus.

- quatre gènes de l'immunité codant pour des peptides impliqués dans la modulation du signal (amplification d'un signal "danger", ou atténuation d'une "fausse alerte"). **Sp14A** active la prophénoloxydase et **Sp14D1** active des ligands SPAETZLE-like intervenant dans le signal de transduction de peptides anti-microbiens. Ces deux sérines protéases sont induites par un challenge à *Plasmodium berghei* chez *An. gambiae* [10]. **CLIP B15** (une autre sérine protéase) présente une réponse suite à une infection à *Plasmodium* semblable à celle de TEP4, tandis que **SRPN 10** montre une induction brève et intense à 28 heures post-infection, période à laquelle les ookinètes traversent l'épithélium intestinal [9]. Ces 4 gènes ont été retenus.

2.1.2. Recherche des séquences nucléotidiques

Elles ont été extraites à partir de la banque de données du génome d'*An. gambiae* (http://ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=agambiae.inf). Les numéros d'accès correspondant à la séquence peptidique de PEST pour les gènes choisis ont été recherchés soit à partir de la littérature, soit à partir de GeneBank. Le numéro d'accès permet de connaître la position de la séquence nucléotidique correspondante au gène sur son scaffold. Un scaffold représente la séquence nucléotidique d'un fragment de chromosome, sa taille peut aller jusqu'à 1 Mb. La séquence nucléotidique recherchée a alors été extraite du scaffold, et l'emplacement des régions non-codantes (introns, 5'-UTR, 3'-UTR) a été notifié.

2.1.3. Détermination des amorces PCR (tableau 1)

Des amorces PCR ont été recherchées de manière à amplifier des fragments de 450 à 1200 pb, selon la taille du gène, et comportant des régions codantes et non codantes. La détermination des amorces a été réalisée à l'aide du logiciel Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

Les critères de choix ont été définis de manière à respecter un pourcentage de G-C d'environ 50%, et un minimum d'hybridation entre les amorces d'un couple (amorces-F sur amorce-R) et entre les exemplaires d'une même amorce.

Pour chaque couple d'amorces, la température d'hybridation optimale a été déterminée par une PCR réalisée dans les conditions standards (tableau 2) pour un gradient de température allant de 50°C à 64°C.

2.2. MATERIEL BIOLOGIQUE

Des moustiques utilisés dans cette étude étaient issus de populations naturelles d'*Anopheles gambiae s.l.* de trois pays africains : Sénégal, Burkina Faso et Cameroun. Les moustiques avaient été capturés :

- pour le Sénégal : à Dielmo, situé à 200 kilomètres au sud est de Dakar, en zone soudanienne, présentant une végétation de type savane arborée.
- pour le Burkina Faso : dans la vallée du Kou située à une trentaine de kilomètres au Nord de Bobo Dioulasso, zones de cultures du riz et du coton. Le régime climatique et la végétation sont du même type que ceux de Dielmo.
- Pour le Cameroun : au village de Simbock, situé près de Yaoundé. Le climat y est de type équatorial et des forêts dégradées constituent le biotope.

Dans chacune de ces localités, *An. gambiae s.s.* et *An. arabiensis* coexistent et les formes moléculaires M et S d'*An. gambiae s.s.* sont trouvés en sympatrie. L'espèce puis la forme moléculaire ont été déterminées en effectuant successivement une PCR diagnostique d'espèce pour la différenciation d'*An. gambiae s.s./An. arabiensis* [13], et une PCR de forme moléculaire (M/S) chez *An. gambiae s.s.* [14]. Les populations d'*An. gambiae s.l.* utilisées comptaient environ 100 moustiques par localité, sauf pour le Burkina Faso où les moustiques avaient déjà été déterminés (espèce et forme moléculaire) au cours d'un travail de doctorat [39]. Les amplifications PCR pour l'étude de la diversité nucléotidique ont été ensuite réalisées sur un nombre de moustiques suffisant pour obtenir les séquences amplifiées de trois individus par localité et par forme moléculaire.

2.3. METHODE

2.3.1. Extraction de l'ADN

Les extractions ont été réalisées à partir du corps entier du moustique femelle adulte selon un protocole adapté de celui de Cornel et Collins [15], et utilisé en routine au LIN. Brièvement, le moustique est broyé à l'aide d'un piston stérile dans 100µL de tampon de broyage (10mM Tris-HCl, 100mM NaCl, 25 mM EDTA, 0,5% SDS, Proteinase K 60 µg/mL) dans un tube stérile de 1,5 mL et incubé pendant 60 min à 55°C afin de lyser les protéines.

Les débris cellulaires sont précipités par addition de 13µL d'acétate de potassium (8M) et la réaction est activée en maintenant sur glace pendant 30 min. Une centrifugation à 4°C à 14000 trs/min pendant 15 min permet de récupérer l'ADN solubilisé dans la phase aqueuse. L'ADN est alors précipité par addition de 200 µL d'éthanol glacial, puis culoté par centrifugation. Le culot d'ADN est ensuite lavé dans 100 µL d'éthanol 70%, puis séché au Speed-Vac et resuspendu dans 20 µL d'eau stérile.

La concentration de l'extrait d'ADN a été contrôlée par spectrophotométrie. Les extraits avaient à une concentration d'environ 400 µg/mL. Les réactions PCR ont été réalisées à partir de dilutions au centième.

2.3.2. Amplification des séquences des gènes par PCR

Les conditions standards de PCR décrites tableau 2 ont été utilisées pour réaliser les amplifications de chaque séquence, dans un volume final de 50µL, selon le cycle d'amplification indiqué tableau 3.

Tableau 2 : conditions standards utilisées pour les PCR

Réactifs	Concentration finale
Tampon d'enzyme 10X	1X
25 mM MgCl ₂	1,5 mM
5mM dNTP	0,2 mM pour chaque dNTP
Amorce-F	10 pmoles
Amorce-R	10 pmoles
Taq polymérase (5U/µL)	1 U
ADN produit d'extraction dilué	Environ 10 ng
Eau stérile	QSP 50µL

Tableau 3 : cycle d'amplification

Etape du cycle	Température du thermocycleur	Temps
Dénaturation initiale	94°C	3min
35 cycles d'amplification		
- dénaturation	94°C	30 sec
- hybridation	Température optimale, spécifique pour chaque couple d'amorce	30 sec
- extension	72°C	45 sec
Extension finale	72°C	5 min

2.3.3. Electrophorèse sur gel d'agarose

La présence, la taille et l'intensité des amplicons ont été évaluées par migration de 5 µL de produit PCR sur gel d'agarose 1% (1g d'agarose dans 100 mL de TBE 1X contenant 0,5 µg/mL de bromure d'éthidium). La migration est réalisée dans du tampon TBE 1X à 250mA, à 140 V pendant 45 minutes. La révélation s'effectue sous UV. Chaque produit de PCR est évalué qualitativement et quantitativement par rapport au marqueur de taille *SmartLadder Small Fragments* (QIAGEN).

2.3.4. Séquençage

Le séquençage des produits de PCR a été sous-traité par Génome Express (Grenoble, France). La compagnie effectue du séquençage automatisé sur séquenceur ABI 3700. Le principe, basé sur la méthode de Sanger, consiste à effectuer une réaction PCR à l'aide de didéoxynucléotides marqués avec 4 fluorochromes différents (Dye-terminator). A la fin de la réaction, les fragments de PCR, de tailles et de couleurs différentes, sont purifiés puis séparés sur gel de polyacrylamide. La lecture de la séquence est réalisée par un scanner au fur et à mesure de la migration. Les séquences nous ont été fournies sous forme d'électrophorégrammes.

2.3.5. Analyse des données

Les électrophorégrammes ont été visualisés à l'aide du logiciel Chromas, Version 1.45 [16]. Ils ont été corrigés manuellement afin d'éliminer les erreurs de lecture et d'identifier les SNPs. La nomenclature décrite par Black a été adoptée pour la codification des bases hétérozygotes [44].

L'alignement des séquences corrigées a été réalisé à l'aide du logiciel ClustalX (1.81) [17]. La taille des fragments à comparer a été ajustée aux plus courtes séquences aux extrémités 5' et 3'. Les alignements ont été inspectés visuellement et corrigés si nécessaire.

L'analyse des séquences a été effectuée à l'aide du logiciel MEGA version 2.1 [18]. Le logiciel permet de calculer, pour chaque gène, la composition en bases nucléotidiques de la séquence. La composition en bases des régions codantes a été reportée pour l'ensemble de la séquence, ainsi que pour les bases situées en troisième position du codon. Le contenu en bases à cette position fournit une indication du biais de l'usage des codons. L'usage des codons a été mesuré en régions codantes pour chaque gène, puis compilé pour l'ensemble des gènes étudiés. Pour cette mesure, un individu de forme moléculaire M du Burkina Faso a été pris comme référence. L'usage des codons donne la répartition des différents codons synonymes

pour chaque acide aminé et permet de mesurer l'usage relatif des codons synonymes (RCSU). Le RCSU correspond, pour un codon donné, au rapport de sa fréquence observée dans la séquence à sa fréquence attendue sous l'hypothèse d'un usage égal des codons [21]. Il mesure ainsi le biais dans l'usage des codons. Un autre indice du biais de l'usage des codons a également été calculé : l'ENC ou nombre effectif de codons [22]. L'ENC est la mesure du nombre effectif de codons utilisés dans un gène : il varie de 20 lorsqu'un seul codon est utilisé pour chaque acide aminé à 61 quand tous les codons sont employés. Il a été déterminé pour chaque gène en utilisant le logiciel Chips du programme EMBOSS (<http://bioinfo.pbi.nrc.ca:8090/cgi-bin/emboss>).

Le dénombrement et la caractérisation des SNPs et insertion/délétion ont été réalisés à l'aide de MEGA et des alignements obtenus sous ClustalX. Les régions codantes et non codantes ont été traitées séparément pour chacun des gènes. En régions codantes, on a distingué les substitutions synonymes (ou silencieuses), intéressant un changement de base qui ne modifie pas l'acide aminé correspondant, des substitutions non synonymes (ou de remplacement) induisant un changement de l'acide aminé. Pour chaque catégorie de substitutions, les SNP ont été classés en transitions (substitutions T↔C, et A↔G,) et transversions (A↔T, C↔G, C↔A, T↔G).

La diversité nucléotidique (π) représente le nombre moyen de substitutions nucléotidiques par site entre deux séquences [19]. Elle a été calculée selon la formule suivante

$$\hat{\pi} = \sum_{i < j}^m d_{ij} / c$$

où m est le nombre total de séquences examinées, c le nombre total de comparaisons effectuées deux à deux, et d_{ij} le nombre de substitutions nucléotidiques par site entre la séquence i et j.

d_{ij} a été estimé selon le modèle de Kimura à 2 paramètres qui prend en compte l'inégale distribution des transitions et transversions [20]. π a été calculée en régions codantes et non codantes pour chacun des gènes.

3. RESULTATS

3.1. AMPLIFICATION ET SEQUENCAGE DES GENES CIBLES

Parmi les 16 gènes initialement choisis, 8 ont pu être analysés pour l'ensemble des populations. La description de ces 8 gènes est donnée dans le tableau 1. Pour les 8 autres

gènes, l'analyse n'a pas pu être réalisée du fait de problèmes rencontrés soit lors de l'amplification, soit au cours du séquençage.

3.1.1. Amplification des séquences

Pour les gènes TEP4, arrestin 1, OR1, les amorces PCR définies à l'aide des séquences de la souche PEST, ont généré des signaux d'amplification trop faibles, voir absents, sur les populations naturelles. Afin d'obtenir de meilleurs rendements, les conditions d'amplification ont été modifiées soit en augmentant le nombre de cycles jusqu'à 40, soit en augmentant la concentration de $MgCl_2$ (2,5 mM), soit en combinant les deux. Néanmoins, peut-être en raison d'un polymorphisme au niveau de la zone d'hybridation des amorces, aucun produit de PCR n'a été obtenu en quantité suffisante, et de nouvelles amorces ont été redéfinies.

Pour les gènes TEP3 et arrestin 1, le rendement d'amplification était très différent selon la provenance des moustiques. Ceci nous a conduit à procéder au clonage de certains individus pour lesquels la réaction PCR ne fournissait pas assez de matériel (4 individus du Cameroun pour TEP 3, et 5 du Cameroun pour arrestin 1). Le clonage a été réalisé à l'aide des kit *Zero Blunt TOPO PCR* (INVITROGEN) selon le protocole fourni. Les plasmides ont ensuite été préparés à l'aide du kit *QIAprep Miniprep Handbook* (QIAGEN) en suivant les recommandations fournies par la compagnie.

3.1.2. Séquençage.

La plupart des produits de PCR ont été séquencés à partir des deux extrémités (sense et antisense) afin d'obtenir une lecture fiable. En raison de problèmes inhérents au séquençage (bruit de fond, 50 premières paires de bases(pb) absentes ou ambiguës...) la taille de l'alignement nucléotidique analysable pour chaque gène est inférieure à la taille du fragment amplifié (tableau 1).

Nous avons été confronté à des individus qui présentaient du polymorphisme d'insertion/délétion à l'état hétérozygote, ce qui rend la lecture impossible à partir de la mutation. Le problème a été résolu en confrontant les résultats des séquences senses et antisenses pour le même individu.

Pour les gènes TEP4, AgMuc1, SRPN10 et CLIP B15, peu de séquences utilisables ont été obtenues du fait de la présence de doubles "clones". Ces doubles clones peuvent être dus à une duplication des gènes dans le génome, ce qui reste peu probable car les gènes choisis étaient uniques dans le génome de la souche PEST, ou à une différence trop importante entre les deux allèles. L'analyse de ces gènes n'a pas été poursuivie.

3.2. BIAIS DANS L'USAGE DES CODONS

Pour l'usage des codons (tableau 4), on observe une préférence pour les triplets se terminant en G ou C pour tous les acides aminés. Le biais vers G ou C en position terminale des codons est particulièrement élevé pour les codons dégénérés 6 fois, avec une valeur de RCSU atteignant 3,13 pour le codon CTG. La préférence pour le G terminal domine celle pour le C terminal, mais elle varie suivant les acides aminés. Par exemple, le biais se dirige majoritairement vers G pour les codons correspondant à V, A, P, L, S mais vers C pour G et R.

L'analyse de la composition en base (tableau 5) montre également la préférence pour les triplets en G ou C. En effet, si on observe des valeurs proches dans l'utilisation des 4 bases quand les séquences sont considérées entièrement, les pourcentages moyens de G et C sont plus élevés que ceux de A et T (36,5% et 34,4% contre 12,5% et 16,5% respectivement) lorsqu'on ne considère que le troisième nucléotide des codons. La tendance pour la préférence de G est même accentuée lorsqu'on ne considère que les troisièmes nucléotides des codons dégénérés quatre fois où G prend une valeur maximale de 39,6% contre 30,2% pour C.

Enfin, nos résultats montrent un ENC variable suivant les gènes, de 36,67 pour Sp14D à 61 pour TEP 3 avec une moyenne de 47,34. Le pourcentage de G et C à la troisième position des codons (GC3) est corrélé négativement à l'ENC ($r = -0,94$; $p < 0,01$).

3.3. POLYMORPHISME NUCLEOTIDIQUE

Un total de 5222 pb a été examiné sur 8 gènes analysés, dont 3902 pb en régions codantes et 1320 en régions non codantes (introns). 275 SNPs ont été identifiés : 173 en régions codantes (dont 4 tri-alléliques) et 102 en régions non codantes (dont 3 tri-alléliques). Trois polymorphismes d'insertion/délétion ont été identifiés en régions non-codantes (tableau 6).

En régions codantes, on observe une inégale répartition des substitutions synonymes et non synonymes. La proportion moyenne de substitutions non synonymes est de 22% : elle varie selon les gènes, allant de 13% pour laminin à 46% pour TEP3. La répartition Transition/Transversion (Ts/Tv) est d'environ $\frac{3}{4}$ - $\frac{1}{4}$ (120-45), et cette répartition est conservée si on considère les substitutions synonymes et non synonymes séparément.

La diversité nucléotidique (π) est très variable suivant les gènes et les régions codantes/non codantes (tableau 7). Elle est minimale pour RpS11 dans le codant ($\pi = 0,0003$) et OR1 dans le non codant ($\pi = 0,0025$), et maximale pour Sp14A dans le codant ($\pi = 0,00119$) et RpL31 dans le non codant ($\pi = 0,0091$). On observe une moyenne de 6,25

substitutions par kilobase dans les régions codantes, et de 29,5 pour les régions non codantes, mais cette différence n'est pas significative. Par contre, le polymorphisme nucléotidique est significativement plus élevé pour les gènes de l'immunité que pour les gènes du comportement ou les protéines ribosomales ($p < 0,001$). Enfin, l'analyse des diversités n'a pas permis de mettre en évidence un effet structurant de populations, relatif à la forme moléculaire ou à l'origine des populations. Cependant, pour les moustiques du Burkina Faso, nous avons mis en évidence, chez TEP3, une mutation non synonyme fixée entre M et S (Y443H). Afin de vérifier que cette observation n'était pas liée à un biais d'échantillonnage, les séquences de TEP3 de 15 individus supplémentaires du Burkina Faso (9 M, 6 S) ont été obtenues. La mutation est restée fixée sur l'ensemble des 31 moustiques du Burkina Faso. De même, au Sénégal, pour le gène Sp14D1, 4 mutations (3 SNPs, 1 indel) ont été identifiées en régions non codantes

4. DISCUSSION

4.1. BIAIS DANS L'USAGE DES CODONS

Les analyses de l'usage des codons et de la composition en bases à la troisième position des codons montrent un biais pour les triplets en G et C. Le biais est particulièrement marqué pour une préférence du G. Mais il varie selon le gène et selon l'acide aminé considéré. Ainsi, les protéines ribosomiales présentent une préférence pour l'usage du C, et l'acide aspartique ne montre pas de biais entre l'utilisation du C ou T. Le biais vers l'usage des codons en G et C a déjà été reporté chez la drosophile et *Aedes*, pour lesquels la préférence est pour le C, et nos résultats confirment les tendances observées par Besansky chez *An. gambiae* [23, 24, 25, 26]. Le biais dans l'usage des codons résulterait de la sélection pour une traduction efficace: l'usage des codons est biaisé vers les codons "optimaux" correspondant aux ARNt les plus abondants.[26,30,31]

On observe une corrélation négative entre l'ENC et le pourcentage de GC3 : ainsi les proportions de GC3 les plus élevées sont trouvées pour les gènes les plus biaisés. Les gènes les plus biaisés, ayant les ENC les plus faibles, correspondent généralement aux gènes soumis à de fortes contraintes fonctionnelles, tels que les facteurs de transcription et les protéines ribosomiales, d'expression ubiquitaire. On trouve en effet un biais très important pour RpL31 (37,94). Le faible biais observé pour RpS11 peut être dû au petit nombre de bases analysées pour la région codante (176). Le gène TEP3, par contre, ne montre aucun biais, ni dans l'usage

des codons, ni dans le nombre effectif de codons utilisés. Il pourrait correspondre à un gène exprimé spécifiquement dans un tissu donné ou induit dans le cadre d'une réponse spécifique. En effet, les biais d'usage des codons sont corrélés au niveau d'expression des gènes : c'est à dire que les gènes de maintenance, fortement exprimés tendent à avoir un biais dans l'usage des codons élevé, alors que pour les gènes peu exprimés, le biais est faible ou nul [21,30].

4.2. POLYMORPHISME NUCLEOTIDIQUE

Au cours de cette étude, nous avons pu observer un fort polymorphisme nucléotidique dans les populations naturelles d'*An. gambiae*. En effet, nos résultats révèlent en moyenne 1 SNP toutes les 59 bases. Nous avons également identifié des indels dans les régions non-codantes de 3 des 8 gènes examinés et ce nombre est sous-estimé puisque certains gènes ciblés (Mucin, GGBP A1, TEP4) n'ont pas pu être exploités car ils présentaient des indels à l'état hétérozygote. Les polymorphismes d'insertion/délétion sont donc relativement fréquents et ils représentent une contrainte pour l'identification de mutations ponctuelles par séquençage direct.

Afin de mieux comprendre le polymorphisme nucléotidique chez *An. gambiae*, nous avons étudié les SNPs en terme de substitutions synonyme/non-synonyme et de mutations de type transition/transversion. Sur l'ensemble des régions codantes examinées les sites polymorphes synonymes représentent 78% des SNPs, ce qui correspond à la proportion généralement observée chez d'autres organismes [25,32]. Les mutations synonymes, n'entraînant pas de modification de la séquence peptidique, sont généralement considérées comme étant neutres. Ce type de substitutions est couramment utilisé pour estimer les paramètres de l'évolution moléculaire, tels que le taux de mutations ou la dérive génétique. Cependant, la préférence dans l'usage des codons pour les triplets en CG₃ doit introduire un biais dans ces estimations [28]. Au contraire des mutations synonymes, les substitutions de remplacement se produisent généralement au niveau de sites soumis à sélection. Le ratio de substitutions synonymes/non-synonymes constitue ainsi une indication de la sélection qui s'opère sur les séquences nucléotidiques. Sur les 8 gènes que nous avons examinés, nous observons une proportion de substitutions non-synonymes significativement plus élevée ($\chi^2=17.09$, $p < 0.01$) pour le gène TEP3 (54% de nsSNPs), ce qui indique qu'un phénomène de sélection positive doit agir sur ce gène de l'immunité.

En régions codantes, la proportion transition/transversion observée chez *An. gambiae*, de l'ordre $\frac{3}{4}$ - $\frac{1}{4}$, correspond aux ratios déjà observés chez l'Homme, la drosophile ou *Aedes aegypti* [25, 32, 33]. Ce déséquilibre est à rattacher à la redondance du code génétique et aux

contraintes sélectives pour la conservation de la structure protéique. En effet, la majorité des SNPs correspondent à des bases situées en troisième position du codon et les transitions, pour ces bases, ne modifient pas l'acide aminé correspondant, à l'exception des 2 acides aminés uniques. Les proportions transition/transversion restent similaires lorsque l'on considère les substitutions synonymes et non-synonymes séparément ($p > 0.5$, ns), ce qui est quelque peu surprenant. Les substitutions non-synonymes induisent un changement de l'acide aminé et il ne devrait plus y avoir, a priori, de sélection en faveur des transitions. On peut alors penser que le niveau plus élevé qu'attendu de transitions est aussi lié aux réactions de désamination des cytosines, qui se produisent fréquemment [33]. Ce phénomène est associé à la méthylation de l'ADN, facteur important dans la régulation de l'expression des gènes ; la méthylation se produit essentiellement au niveau des résidus cytosine et engendre une désamination spontanée aboutissant à une thymine.

Les diversités nucléotidiques, définies comme le nombre moyen de substitutions nucléotidiques par site entre 2 séquences, ont été calculées en régions codantes et non-codantes pour les 8 gènes examinés. Nous observons que les diversités ne sont pas significativement différentes selon ces deux régions ($p = 0.103$). Ceci pourrait être dû au plus petit nombre de bases non-codantes (1320 contre 3902 bases codantes). Cependant, chez la drosophile, on soupçonne que les introns et les régions flanquantes des gènes sont également soumis à sélection, probablement du fait de contraintes d'ordre structural et fonctionnel [32,27]. Nos résultats montreraient alors que ces contraintes sélectives sur les régions non-codantes existent aussi chez *An. gambiae*. Les diversités en régions codantes et non-codantes des gènes apparaissent corrélées chez la drosophile et *Ae. aegypti* [25,32] et cela pourrait refléter que le degré de polymorphisme est sous le contrôle de facteurs régionaux, tel que le taux de recombinaison. Chez *An. gambiae*, nous retrouvons également cette corrélation, sauf pour le gène RpL31. Ce gène présente une diversité nucléotidique très élevée en région non-codante en raison de 15 mutations successives se produisant sur une séquence 32 de bp. Si cette séquence est considérée comme un seul événement mutationnel, la corrélation est rétablie.

Les diversités nucléotidiques varient selon le gène considéré, allant de 0.0003 à 0.0119 pour les gènes RpS11 et Sp14A, respectivement, en région codante et de 0.00925 à 0.0519 pour Or1 et SP14A en région non-codante. On observe des différences selon la catégorie fonctionnelle des gènes. Ainsi, les gènes impliqués dans la réponse immunitaire du moustique présentent des diversités significativement plus élevées que les gènes codant pour les protéines ribosomiales et les gènes de l'olfaction ($p < 0.01$). Ces différences sont à relier aux

contraintes sélectives agissant sur les gènes ; en effet, les gènes de maintenance subissent de fortes pressions de sélection pour la conservation de la protéine alors que les gènes impliqués dans une adaptation spécifique seront sujet à une sélection positive favorisant la diversité au niveau des acides aminés [34,36]. Les gènes de l'olfaction OR1 et Arrestin, peu polymorphes, seraient ainsi des gènes conservés. Ces gènes peuvent être impliqués dans des comportements de recherche de l'hôte nourricier ou du partenaire sexuel et il serait alors intéressant de mesurer la divergence interspécifique avec d'autres espèces anophéliennes anthropophiles et zoophiles.

Les diversités nucléotidiques importantes trouvées chez les gènes de l'immunité, SP14A, SP14D, Laminin et TEP3, reflètent certainement une sélection adaptative. Les interactions hôtes-pathogènes entraînent une «course aux armements»: le pathogène évolue pour échapper aux réactions de défense de son hôte tandis que le système de défense de l'hôte évolue de manière à contrecarrer ces phénomènes d'évasion. Les gènes impliqués dans ces interactions doivent alors pouvoir évoluer rapidement. Et en effet, chez la drosophile, il a été montré qu'une évolution adaptative se produisait au niveau du domaine I κ B du gène de l'immunité Relish [29]. De même, une évolution rapide est observée dans les gènes codant pour des immunoglobulines chez l'homme [36] et les protéines de surface de *Plasmodium falciparum* présentent un polymorphisme particulièrement élevé [38]. Nos résultats pourraient alors indiquer que les 4 gènes de l'immunité que nous avons étudiés sont impliqués dans des interactions vecteurs-pathogènes, d'autant que nos moustiques provenaient de zones d'endémie palustre.

Le gène TEP3 apparaît particulièrement intéressant : 1/ il ne présente pas de biais dans l'usage des codons, ce qui peut traduire qu'il évolue très rapidement ou qu'il a une expression spécifique 2/ il a une proportion de substitutions de remplacement élevée ce qui indiquerait une sélection positive et 3/ il exhibe une substitution non-synonyme fixée entre les formes moléculaires M et S du Burkina Faso (mutation Y433H). Selon la théorie neutre de l'évolution moléculaire, la majorité des mutations fixées est due à une dérive aléatoire de changements nucléotidiques sélectivement neutres. Cependant, il est peu probable que la mutation Y443H de TEP3, fixée entre les formes moléculaires M et S du Burkina Faso, soit liée à une dérive génétique. En effet, si une barrière pré-copulatoire isole en zone de sympatrie les formes M et S, des échanges géniques existent entre les différentes populations appartenant à la même forme (M ou S) de l'Est à l'Ouest de l'Afrique. La possibilité d'un biais d'échantillonnage pourrait être soulevée, mais dans ce cas des SNPs fixés seraient également observés pour les autres gènes et cette hypothèse peut donc être écartée. Cette mutation fixée

Y443H de TEP3 serait alors plutôt liée à la sélection naturelle, et il doit exister dans cette localité du Burkina Faso des pressions de sélection distinctes entre les 2 formes moléculaires M et S. La sélection a entraîné la fixation de cette substitution de remplacement, ce qui indique une valeur adaptative importante. Dans la Vallée du Kou, contrairement à Simbock et à Dielmo, les gîtes larvaires des formes moléculaires M et S sont différents, les gîtes des formes M se trouvant dans les rizicultures et ceux des formes S dans des flaques temporaires [39]. TEP3 est associé à la reconnaissance d'agents pathogènes et l'expression du gène est induite suite à une exposition bactérienne [9]. Il est alors possible que la présence d'une bactérie entomo-pathogène dans les gîtes larvaires de l'une des formes moléculaires d'*An. gambiae* ait conduit à une différenciation moléculaire. Cette hypothèse devra être vérifiée en génotypant la mutation à plus large échelle, sur de nouvelles populations d'*An. gambiae*. Des analyses complémentaires devront aussi être effectuées pour SP14D1, afin de vérifier si les 4 mutations sont fixées dans les populations des formes M et S de Dielmo, ce qui sous-tendrait là encore une adaptation localité-dépendante.

5. CONCLUSION

L'étude de la diversité nucléotidique de populations naturelles d'*Anopheles gambiae* s.s. a permis de montrer un polymorphisme important chez 4 gènes impliqués dans la réponse immunitaire : Sp14D1, Sp14A, laminin et TEP3. Ces gènes pourraient être soumis à une sélection adaptative en zone d'endémie parasitaire. Des études d'associations entre le polymorphisme génétique de ces gènes et la permissivité des moustiques au *Plasmodium* seraient intéressantes à mener, en vue d'identifier des locus candidats associés au niveau d'infection parasitaire du moustique. Les gènes OR1 et arrestin, impliqués dans l'olfaction ont présenté peu de polymorphisme, probablement du fait que les trois populations étudiées sont anthropophiles.

Le champ d'études ouvert grâce au génome d'*Anopheles gambiae* fournira indéniablement une meilleure connaissance des mécanismes impliqués dans des interactions *Plasmodium-Anopheles*, et permettra une meilleure compréhension de la compétence et de la capacité vectorielle du moustique. Néanmoins, l'utilisation de ces nouvelles connaissances pour la lutte ("production" de moustiques transgéniques ne pouvant plus transmettre le parasite) sera beaucoup plus délicate et longue. La lutte anti-vectorielle actuelle devrait aussi s'organiser autour d'une meilleure utilisation des moyens existants déjà (insecticides, moustiquaires....).

Tableau 1 : description des gènes utilisés

Fonction	Gène	Chro.	Localisation	N°accession	Scaffold	Amorces	Ind.	PCR	Align.
Immunité : reconnaissance	TEP 3	3	39C-40A	EAA10529	AAAB01008951.1	F : GTGGTGAAGGAACAGGCAAT R : TTACCTCCCACAATCTTGC	15	1192	782
Immunité : modulation du signal	Sp 14A (=CLIP B9)	2	14A-14C	EAA08418	AAAB01008879.1	F : CCGTTGCATCTCCATCTACG R : ATGCCCAGCTTTTGCTTAAC	18	985	714
	Sp 14D1 (=CLIP B4)	2	14C-14D	EAA45573.1	AAAB01008794.1	F : ACGAACGCTACATCCTCACC R : AGTCGCTCGTTACAGGTCGT	17	816	682
récepteurs constitutifs de l'intestin	LAMININ	2	25D-28D	EAA04042	AAAB0100887.1	F : AAGATCAGCAAGGGTGGCTA R : CGTACTTGTCTGGGCTGTT	18	1100	551
olfaction	ARRESTIN 1	3	34A-37D	EAA13874	AAAB01008980.1	F : CGTTCACGTTCAACATCTCG R : TGCTGGTCGATTGTATCCTG	17	974	728
	OR1	3	34-37D	EAA13838	AAAB01008980.1	F : CTATGGCCACCGGAAGATAC R : TTCGCATCTCTTCCATTCC	17	1091	782
gènes de maintenance	RpS11	3	34A-37D	EAA13929	AAAB01008980.1	F : TTTTCGCACCGTAAAGATGG R : GACAGATGCACGCTCATGTT	18	1031	787
	RpL31	3	43B-44B	EAA00150	AAAB01008986.1	F : AAAGTCGGCGATCAATGAAG R : GATTCGACGTTCTCGGTCTG	18	436	327

Chro. = chromosome, **Ind.** = nombre d'individus séquencés, **PCR**= taille de la séquence amplifiée, **Align.** = taille de l'alignement analysé.

Tableau 6: polymorphisme d'insertion/délétion

gène	nombre	taille (pb)
Sp14D	4	4, 4, 11, 15
RpL31	1	1
RpS11	1	1

NB : ces 6 insertions/délétions ont été identifiées en régions non codantes.

Tableau 4 : usage des codons chez *Anopheles gambiae* s.s.

aa	codon	N	RCSU	aa	codon	N	RCSU	aa	codon	N	RCSU	aa	codon	N	RCSU	aa	codon	N	RCSU						
codons dégénérés deux fois																									
Phe (F)	UUU	16	0.70	Cys (C)	UGU	8	0.70	His (H)	CAU	7	0.50	Asp (D)	GAU	34	0.96	Glu (E)	GAA	27	0.73						
	UUC	30	1.30		UGC	15	1.30		CAC	21	1.50		GAC	37	1.04		GAG	47	1.27						
Tyr (Y)	UAU	8	0.37	Asn (N)	AAU	22	0.72	Gln (Q)	CAA	13	0.45	Lys (K)	AAA	15	0.36	Ile (I)	AUU	19	0.65						
	UAC	35	1.63		AAC	39	1.28		CAG	58	1.55		AAG	68	1.64		AUC	54	1.84						
																	AUA	15	0.51						
codons dégénérés 4 fois																									
Val (V)	GUU	16	0.76	Ala (A)	GCU	14	0.57	Thr (T)	ACU	10	0.48	Pro (P)	CCU	6	0.44	Gly (G)	GGU	25	1.11						
	GUC	20	0.95		GCC	22	0.89		ACC	31	1.48		CCC	14	1.02		GGC	36	1.60						
	GUA	11	0.52		GCA	20	0.81		ACA	12	0.57		CCA	8	0.58		GGA	21	0.93						
	GUG	37	1.76		GCG	43	1.74		ACG	31	1.48		CCG	27	1.96		GGG	8	0.36						
codons dégénérés 6 fois																									
Leu (L)	UUA	3	0.16	Ser (S)	UCU	0	0.00	Arg (R)	CGU	16	1.20	codons uniques				Met (M)	AUG	29	1.00	Trp (W)	UGG	11	1.00		
	UUG	13	0.68		UCC	15	1.23		CGC	38	2.85														
	CUU	12	0.63		UCA	2	0.16		CGA	9	0.68														
	CUC	16	0.83		UCG	32	2.63		CGG	14	1.05														
	CUA	11	0.57		AGU	5	0.41		AGA	1	0.08														
	CUG	60	3.13		AGC	19	1.57		AGG	2	0.15														
																	codons stop								
																	stop	UAA	0	0.00					
																		UAG	2	3.00					
																		UGA	0	0.00					

aa : acide aminé; RCSU : relative synonymous codon usage; N : nombre de codons; Nombre total de codons examinés : 1297

Tableau 5: composition en base chez *Anopheles gambiae* s.s. (en régions codantes).

gène	N	ENC	TOTAL				Troisième nucléotide des codons (1)				3ième nucléotide des codons dégénérés 4 fois				
			T	C	A	G	T	C	A	G	nombre	T	C	A	G
Sp14D	597	36.67	18.4	27.5	21.9	32.2	14.1	39.2	5.5	41.2	97	10.3	33	5.2	51.5
Sp14A	635	49.21	17.5	28.5	21.9	32.1	16.1	35.5	13.7	34.6	115	16.5	31.3	14.8	37.4
Laminin	551	41.41	13.8	26.3	28.7	31.2	14.2	29.5	11.5	44.8	84	10.7	25	11.9	52.4
RpL31	224	37.94	16.1	33	23.2	27.7	10.8	48.6	6.8	33.8	39	12.8	48.7	10.3	28.2
RpS11	176	56.72	19.3	26.1	31.8	22.7	17.2	37.9	15.5	29.3	23	21.7	30.4	21.7	26.1
Arrestin 1	638	43.14	19.6	28.1	22.6	29.8	14.2	34.4	11.3	40.1	111	14.4	32.4	15.3	37.8
OR1	589	52.63	28	24.3	23.8	23.9	22.4	30.6	16.3	30.6	88	20.5	22.7	19.3	37.5
TEP3	492	61	23.6	23.8	29.5	23.2	22.6	26.8	23.2	27.4	79	20.3	26.6	24.1	29.1
Total	3902										636				
Moyenne		47.34	19.8	27.9	24.7	28.6	16.5	34.4	12.5	36.5		15.4	30.2	14.8	39.6

N: taille de la séquence en pb; ENC: nombre effectif de codons, (1): pour l'ensemble de la séquence codante. Les fréquences nucléotidiques sont données en pourcentage.

Tableau 7 : diversité nucléotidique

	Séquences codantes										Séquences non codantes		
	Taille		Substitutions synonymes			Substitutions non synonymes			Diversité nucléotidique		Taille	Diversité nucléotidique	
	pb	aa	Ts	Tv	Total	Ts	Tv	Total	π	s.e.	pb	π	s.e.
Sp14D	597	199	10	10	20	3	1	4	0.0088	0.0021	85	0.0318	0.0123
Sp14A	635	211	27	7	34	6	3	9	0.0119	0.0025	79	0.0519	0.0152
Laminin	551	183	25	8	34*	3	2	5	0.0104	0.0020	0		
RpL31	224	74	5	3	9*	2	0	2	0.0054	0.0018	103	0.0910	0.0233
RpS11	176	58	1	0	1	0	0	0	0.0003	0.0003	611	0.0120	0.0023
Arrestin 1	638	212	11	2	13	2	1	3	0.0036	0.0010	90	0.0096	0.0061
OR1	589	196	5	4	9	2	0	2	0.0033	0.0012	193	0.0025	0.0017
TEP3	492	164	8	2	11*	10	2	13*	0.0063	0.0015	159	0.0083	0.0029
Total	3902	1297	92	36	131	28	9	38			1320		
Moyenne									0.0062			0.0295	

pb : paire de base; aa : acide aminé; Ts : transition; Tv : transversion; π : diversité nucléotidique; s.e. : standard error; * : un SNP tri-allélique inclus

PROJET DE RECHERCHE

1. INTITULE

Caractérisation du risque de transmission de la Fièvre de la Vallée du Rift par *Aedes vexans*.

2. CONTEXTE

2.1. Présentation de la maladie

La Fièvre de la Vallée du Rift (FVR), considérée jusqu'aux années 60 comme une maladie relativement secondaire a suscité l'intérêt des vétérinaires et médecins en raison de flambées épizoo-épidémiques de plus en plus nombreuses ces dernières années. C'est une arbovirose (liste A de l'OIE) commune à l'homme et à de nombreuses espèces animales. Elle est due à un virus de la famille des *Bunyaviridae*, genre *Phlebovirus*. Elle entraîne le plus souvent, chez l'homme, un simple syndrome grippal qui peut parfois s'aggraver en différentes formes, notamment la forme hémorragique qui atteint un taux de létalité de 50%. Les petits ruminants, puis les bovins, sont les espèces animales les plus sensibles : la maladie se manifeste essentiellement par de la mortinatalité et des avortements (touchant jusqu'à 90% du troupeau).

D'abord limitée à l'Afrique de l'Est et du Sud (épizootie en Afrique du Sud en 1951, au Zimbabwe en 1978), la maladie s'est propagée en Egypte (épidémie en 1977), puis à l'Afrique de l'Ouest (épidémie en 1988 à Rosso en Mauritanie et 1998 dans le sud ouest mauritanien), et a frappé la péninsule arabique en 2000 [53].

2.2. Importance

Malgré des épizooties restant malgré tout rares et dispersées temporellement et géographiquement, l'importance économique de la FVR est considérable (100 000 moutons tués en Afrique de Sud), d'autant plus qu'elle touche des populations pour lesquelles l'élevage est la seule ressource [49]. La multiplication des cas humains à partir de 1977 (600 cas fatals en Egypte, 224 à Rosso) souligne l'importance sanitaire de cette maladie [50]. En période inter-épidémique, les nombreuses enquêtes sérologiques montrent la circulation à bas bruit du virus dans différentes régions, ce qui reflète son expansion insidieuse, et pose la question de

l'émergence potentielle de foyers épidémiques, à partir de ces zones endémiques [51,52]. Ces éléments conduisent à considérer la FVR comme une maladie émergente préoccupante.

2.3. Epidémiologie

La transmission de la maladie est essentiellement vectorielle pour les populations animales. Le virus réalise un cycle entre les vertébrés qui développent une simple virémie ou exprime cliniquement la maladie, et le vecteur biologique (un moustique). Les enquêtes sérologiques montrant une circulation virale, sans ou avec peu de symptômes cliniques, reflètent que ce cycle a lieu en périodes inter-épizootiques dans de nombreux pays. Il est probable que l'amplification de ce cycle, par des mécanismes qui restent à déterminer, conduise à une épizootie. Les humains sont contaminés essentiellement par contacts directs avec des vertébrés virémiques (manipulation des avortons...), donc suite à une épizootie.

2.4. Les vecteurs

Différents vecteurs prouvés et potentiels ont été identifiés, en se basant sur leur abondance, leurs préférences trophiques, des isollements de RVFV et parfois des tests de compétence vectorielle. Ces vecteurs sont différents selon les régions : *Aedes cumminsii*, *Ae.circumluteolus* et *Ae.mcintoshi* en Afrique de l'Est et du Sud, *Ae.vexans*, *Ae.ochraceus* et *Culex poicilipes* en Afrique de l'Ouest, *Ae.caspius*, *Cx.pipiens*, et *Cx.perexiguus* en Egypte, et *Ae.vexans* et *Cx.tritaeniorhynchus* pour la péninsule arabique [54,55,56,57,58,59,60,63,64].

2.5. *Aedes vexans*

Cette espèce connaît une répartition très large sur la planète. Elle est présente en Amérique du Nord, en Europe, sur le pourtour méditerranéen, en Afrique sahélienne, en Asie et dans la zone Pacifique [62]. Elle est particulièrement abondante en Amérique du Nord, en Europe et en zone sahélienne, ce qui peut représenter un risque potentiel d'émergence d'épizooties dans ces zones si sa capacité et sa compétence vectorielle sont suffisantes. Ce risque est d'autant plus réel que le développement des échanges commerciaux favorisent l'introduction du virus dans des zones indemnes, que les grands aménagements hydro-agricoles et le réchauffement de la planète peuvent favoriser la pullulation des populations vectorielles et la diminution de la durée du cycle extrinsèque du virus chez le moustique, et que l'augmentation des densités des populations humaines et animales favorise le contact hôtes-vecteurs.

3. OBJECTIF

L'objectif est de caractériser le risque de transmission de la FVR par *Ae. vexans*, dans des zones où ce moustique est présent. Ce projet vise donc à étudier certains éléments de la capacité vectorielle de cette espèce dans ces régions. Une approche reposant sur 3 axes d'études a été développée : étude d'éléments de la biologie, de la compétence vectorielle et de la structuration des populations. Croiser ces 3 axes permettra de répondre à l'objectif fixé, et d'expliquer des éventuelles différences de risque.

4. PROTOCOLE

4.1. Zones étudiées

Ae. vexans est très abondant en Amérique du Nord en Europe et en zone sahélienne. Une localité d'étude sera donc identifiée en Amérique du Nord, une en Europe et trois dans la bande sahélienne : une au Sénégal, une au Tchad, et une autre au Nord de l'Ethiopie, réalisant ainsi un transect sur cette bande. Dans chaque localité, les étapes suivantes seront menées.

4.2. Etude de certains caractères biologiques

Ce projet de recherche vise tout d'abord à déterminer si *Ae. vexans* représente un vecteur potentiel efficace de FVR. Les modalités du contact hôte-vecteur (étude de la dynamique spatiale et temporelle des moustiques) ne pourront être explorées dans un premier temps dans chaque localité. Elles devraient faire l'objet de recherches approfondies si le risque de transmission de FVR est confirmé dans une localité. Pour ces raisons, seuls seront donc observés des caractères statiques de la biologie du moustique.

- ◆ Identification des gîtes larvaires : le but est de déterminer les caractéristiques des gîtes larvaires spécifiques d'*Aedes vexans*. Les différents types de gîtes de la localité seront recensés et caractérisés de manière exhaustive (taille, "forme", configuration, permanent/temporaire, type de végétation, , localisation, milieu urbain/rural...). Les gîtes larvaires sont des caractéristiques robustes d'espèces : une même espèce utilisera, en principe, les mêmes types de gîtes. Ce caractère fournira donc d'éventuels indices sur l'existence de plusieurs espèces cryptiques au sein d'un complexe (cf.infra)
- ◆ Etude des préférences trophiques : un vecteur d'une maladie ne sera efficace que s'il présente une préférence marquée pour les espèces sensibles à cette maladie. Les préférences trophiques seront étudiées selon deux méthodes différentes : en utilisant des pièges à appâts, et en analysant les repas sanguins de moustiques capturés gorgés et au

repos dans la nature. Les pièges utiliseront le pouvoir attracteur de bovins, équidés, petits ruminants, oiseaux, rongeurs sauvages locaux, et humains.

4.3. Etude de la compétence vectorielle.

Lorsqu'un moustique s'infecte en se gorgeant sur un vertébré en phase virémique, le virus subit, dans l'organisme de l'insecte, une phase d'éclipse, puis se réplique avant de gagner les glandes salivaires (ces étapes constituent le cycle extrinsèque). Le moustique devient alors infectant. Pour être un vecteur efficace, le moustique doit permettre au virus de réaliser ce cycle, ce que définit la compétence vectorielle (compatibilité génétique du vecteur et du virus).

Cette étude sera réalisée dans les laboratoires P3 du Cirad-EMVT à Montpellier. En effet, la législation prévoit l'utilisation de laboratoire sécurisés pour manipuler le RVFV. Nous utiliserons des adultes fournis par un élevage créé avec des œufs provenant de chaque localité. La compétence vectorielle des *Ae. vexans* pour le RVFV pourra ainsi être évaluée pour chaque localité, selon un protocole adapté de celui de Turell [59].

4.4. Etude de la systématique et de la structuration des populations d'*Ae. vexans*

Face à une répartition aussi large, on peut émettre l'hypothèse qu' *Ae. vexans* représente en fait un complexe d'espèce. Il est donc indispensable de s'intéresser à la structuration de ce complexe éventuel : ceci permettra de caractériser les différents taxons de ce groupe, et pourrait expliquer des différences de capacité et compétence vectorielle en cas de présence de plusieurs taxons.

- ♦ Etude des séquences ITS1 et ITS2 (Internal noncoding Transcribed Spacer): ce sont des régions nucléotidiques hypervariables séparant respectivement les séquences codant pour les ARNr 18s, 5,8s et 28s. L'intérêt de ces séquences réside dans leur polymorphisme, et leur amplification par réaction PCR en utilisant des amorces consensus définies sur les séquences hyper conservées des ARNr (de ce fait, des amorces déjà décrites pour une espèce voisine ont de grandes chances de s'hybrider). Les ITS1 et ITS2 seront amplifiées (en utilisant des amorces déjà utilisées sur *Ae. vexans*) et séquencées pour des individus de chaque localité (le clonage pourra être utilisé en cas de polymorphisme intra-individuel trop important) [61]. Elles seront ensuite alignées, et les mutations fixées pour une localité seront recherchées. Dans certaines conditions, ces mutations sont des indicateurs de spéciation et elles permettront ainsi de déterminer des PCR diagnostiques de taxons du

complexe *vexans*. Les distances nucléotidiques pourront également être calculées entre les séquences des individus des différentes localités.

- ♦ Réalisation d'une banque de micro satellites : elle permettra d'identifier des micro satellites (séquences répétées d'ADN, dont le nombre de répétitions détermine un allèle) utilisables pour étudier de façon plus fine la structuration du complexe entre les localités (utilisation des Fst).

L'utilisation de cette double approche permettra de compenser des éventuelles difficultés rencontrées dans une des techniques. Ces résultats seront confrontés aux observations réalisées pour les caractères biologiques, et notamment les gîtes larvaires.

5. RESULTATS ATTENDUS

Ces différentes approches permettront de caractériser le risque de transmission de la FVR par *Ae. vexans*, dans des régions où il est présent. Des différences éventuelles pourront être ramenées à une structuration des populations, ou à l'existence de taxons cryptiques, qui ne présentent pas les mêmes comportements, ou la même compétence vectorielle.

6. APPLICATIONS

La découverte d'un risque de transmission permettrait de sensibiliser les différents acteurs à cette maladie potentiellement présente sur leur territoire. Ceci pourrait s'accompagner de réseaux d'épidémiosurveillance de cette maladie, soit par surveillance sérologique de troupeaux sentinelles, soit par recherche du virus sur "pools" d' *Ae.vexans* capturés dans la nature

La découverte d'un risque de transmission justifierait également des études plus fines dans les localités concernées (âge des populations, durée du cycle extrinsèque aux températures locales, dynamique des populations...) pour déterminer s'il s'agit d'un risque actuel ou potentiel (en cas de réchauffement climatique par exemple...).

Enfin, la découverte d'une structuration importante ou d'indices supposant une spéciation au sein du groupe *vexans* entre les trois localités de la bande sahélienne justifierait également une étude plus fine de génétique des populations sur le même transect.

BIBLIOGRAPHIE MEMOIRE DE DEA

1. WHO. Evaluation de la santé. Rapport sur la santé dans le monde. La vie au 21^{ème} siècle, une perspective pour tous. *World Health Organisation (ed)* 1998:43-65.
2. Ayala, F.J. and S. M. Rich, Evolution of *Plasmodium* and the recent origin of the world populations of *Plasmodium falciparum*. *Parasitologia*, 1999. **41**:p. 55-68.
3. Mouchet, J. and P. Carnevale, Les vecteurs et la transmission, in *Paludisme*, ELLIPSE/AUPELF, Editor. 1991
4. Rodhain, F. and C. Perez, *Précis d'entomologie médicale et vétérinaire*, ed. Maloine. 1985, Paris.
5. Fontenille, D. and Lochouart, The complexity of the malaria vectorial system in Africa. *Parasitologia*, 1999. **41** (1-3): p 267-71.
6. Hamon, J., J.P. Adam, and A. Grjebine, Observation sur la répartition et le comportement des anophèles de l'Afrique-équatoriel Française, du Cameroun et de l'Afrique occidentale. *Bull. Org. Mond. Santé*, 1956. **15**: p. 549-591.
7. Trape, J.F. et al., Combating malaria in Africa. *Trends in Parasitology*, 2002. **18**(5): p.224-230.
8. Vlachou, D. et al., *Anopheles gambiae* laminin interacts with the P25 surface protein of *Plasmodium berghei* ookinetes. *Molecular and Biochemical Parasitology*, 2001. **112**: p. 229-237.
9. Christophides, G.K. et al., Immunity-related and genes families in *Anopheles gambiae*. *Science*, 2002. **298**: p.159-165.
10. Gorman, M.J. et al., Molecular characterisation of five serine protease genes cloned from *Anopheles gambiae* hemolymph. *Insect Biochemistry and Molecular Biology*, 2000. **30**: p. 35-46.
11. Fox, A.N. et al., Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proc. Natl. Acad Sci USA*, 2001. **98**: p14693-14697.
12. Merrill, C.E., et al., Visual arrestin in olfactory pathways of *Drosophila* and the malaria vector mosquito *Anopheles gambiae*. *Proc. Natl. Acad Sci USA*, 2002. **99**: p1633-1638.
13. Scott, J.A. et al., Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg*, 1993. **49**(4): 520-9.
14. Favia, G. et al., Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Mol Biol*, 2001. **10**(1): 19-23.
15. Cornel, A.J. and F.H. Coolins. PCR of the ribosomal DNAZ intergenic spacer regions as a method for identifying mosquitoes in the *Anopheles gambiae* complex. *Methods Mol Biol.*, 1996. **50**:321-32.
16. Mac Carthy, C., (1996) Chromas, version 1.45.
17. Thompson, J.D. et al., The ClustalX windows interface : flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acid research*, 1997. **24** : 4876-82.
18. Kumar, S., et al., (2001) MEGA2 : Molecular Evolutionary Genetics Analysis software. Bioinformatics submitted.
19. Nei, M. and W.-H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad Sci USA*, 1979. **76**: 5269-5273.
20. Kimura, M, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111-120.
21. Sharp, P.M. and W.-H. Li, An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.*, 1986. **24**: 28-38.
22. Wright, F., The effective number of codons used in a gene. *Gene*, 1990. **87**:23-9.
23. Besansky, N.J., Codon usage patterns in chromosomal and retrotransposon genes of the mosquito *Anopheles gambiae*. *Insect Mol. Biol.*, 1993. **1**: 171-8.
24. Powel, J.R. and E.N. Moriyama, Evolution of codon usage bias in *Drosophila*. *Proc. Natl. Acad Sci USA*, 1997. **94**: 7784-7790.
25. Morlais, I. and D.W. Severson, Single nucleotide polymorphisms in *Aedes aegypti*. Publication à venir.
26. Moriyama, E.N. and J.R. Powel, Codon usage and tRNA abundance in *Drosophila*. *J. Mol. Evol.*, 1997. **45**:514-523.
27. Mac Vean, G. A. T. and J. Vieira, Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics*, 2001. **157** : 245-257.
28. Comeron, J. M. and M. Aguade, An evaluation of measures of synonymous codons usage bias. *J Mol Evol*, 1998. **47** : 268-74.
29. Begun, D. J. and P. Whitley, Adaptive evolution of relish, a *Drosophila* NF- κ B/I κ B protein. *Genetics*, 2000. **154** : 1231-1238.
30. Ikemura, T., Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*, 1985. **2**: 13-34.
31. Duret, L., tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet*, 2000. **16**: 287-9.

32. Moriyama, E.N. and J.R. Powel, Intraspecific nuclear DNA variation in *Drosophila*. *Mol Biol Evol*, 1996. **13**: 261-77.
33. Brookes, A.J., The essence of SNPs. *Gene*, 1999. **234**: 177-86.
34. Moriyama, E.N. and J.R. Powel, Synonymous substitution rates in *Drosophila* : mitochondrial versus nuclear genes. *J. Mol. Evol*, 1997. **45**: 378-91.
35. Ganières, J.-P., Maladies réputées contagieuses et maladies à déclaration obligatoire des ruminants. 2002.
36. Hughes, A. L., Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. *Mol Biol Evol*, 1997. **14** (1) : 1-5.
37. Ito, J. et al., Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature*, 2002. **417** : 452-455.
38. Hughes, M. K. and A. L. Hughes, Natural selection on *Plasmodium* surface proteins. *Mol Biochem Parasitol*, 1995. **71** : 99-113.
39. Diabate, A, Le paludisme au Burkina faso : étude de la transmission et répartition géographique de la résistance d' *Anopheles gambiae* sl aux pyréthrinoides. Thèse de doctorat, Université Montpellier II, 2003; 123 p.
40. Shen, Z. et al., A cell surface mucin specifically expressed in the midgut of the malaria mosquito *Anopheles gambiae*. *Proc. Natl. Acad Sci USA*, 1999. **96**: 5610-5615.
41. Holt, R. A. et al., The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, 2002. **298** : 129-149.
42. Collins, F. H. et al., Genetic selection of a *Plasmodium* refractory strain of the malaria vector *Anopheles gambiae*. *Science*, 1986. **234** : 607-610.
43. Zheng, L. et al., Quantitative trait loci for refractoriness of *Anopheles gambiae* to *Plasmodium*. *Science*, 1997. **276** : 425-428.
44. Black, W. C., et al., Population genomics : genome wide sampling of insect populations. *Annu Rev Entomol*, 2001. **46** : 441-69.
45. Dimopoulos, G. et al., Molecular immune response of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc. Natl. Acad Sci USA*, 1997. **94** : 11508-11513.
46. Dimopoulos, G. et al., Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle. *The EMBO Journal*, 1998. **17** (21) : 6115-6123.
47. Hugot, J. P. et al., Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, 2001. **411** : 599-603.
48. Dimopoulos, G. et al., Genome expression analysis of *Anopheles gambiae* : responses to injury, bacterial challenge and malaria infection. *Proc. Natl. Acad Sci USA*, 2002. **99**(13) : 8814-19.
49. Gouagna, L.C. et al., The early cycle of *Plasmodium falciparum* in laboratory-infected *Anopheles gambiae* : an estimation of parasite efficacy. *Trop Med Int Health*, 1998. **3** (1) : 21-8.

BIBLIOGRAPHIE PROJET DE RECHERCHE

49. Sculz, K. H., Rift Valley Feverin South Africa. *Special Report 5/51. Union Department of Health*, 1951.
50. Lefevre, P.C., Actualité de la fièvre de la vallée du Rift. Quels enseignements tirer de l'épidémie de 1977 et 1987? *Med Trop*, 1997. **57** : 61-64.
51. Morvan, J. et al., Rift Valley fever in Madagascar in 1991. Sero-epidemiological studies in cattle. *Rev Elev Vet Pays Trop*, 1992 a. **45** (2) : 121-7.
52. Morvan, J. et al., Duration of immunoglobulin M antibodies against Rift Valley fever virus in cattle after natural infection. *Trans R Soc Trop Med Hyg.*, 1992 b. **86** (6) : 675.
53. Jouan, A. et al., An Rift Valley Fever epidemic in southern Mauritania. *Ann Inst Pasteur Virol*, 1988 Jul-Sept. **139** (3): 307-8.
54. Turell, M. J. and C. L. Bailey, Transmission studies in mosquitoes (*Diptera* : *Culicidae*) with disseminated Rift valley Fever virus infections. *J Med Entomol*, 1987. **24** (1) : 8-11.
55. Logan, T. M. and K. J. Linthicum, Eggs hatching of *Aedes* mosquitoes during successive floodings in a Rift Valley fever endemic area in Kenya. *J Am Mosq Control Assoc*, 1991. **7** (1) : 109-112.
56. Fontenille, D. et al., New vectors of Rift Valley fever in West Africa. *Emerg Infect Dis*, 1998. **4** (2) : 289-93.
57. Gad, A. M. et al., Host feeding patterns of *Culex pipiens* and *Cx antennatus* (*Diptera* : *Culicidae*) from a village in Sharqiya Governorate, Egypt. *J Med Entomol*, 1995. **32** (5) : 573-7
58. Turell, M. J. et al., Isolation of west Nile and sindbis viruses from mosquitoes collected in the Nile Valley of Egypt during an outbreak of Rift Valley fever. *J Med Entomol*, 2002. **39** (1) : 248-50.
59. Turell, M. J. et al., Vector competence of Egyptian mosquitoes for Rift Valley fever virus. *Am J Trop Med Hyg*, 1996. **54** (2) : 136-9.

60. Jupp, P. G. et al., The 2000 epidemic of Rift Valley fever in Saudi Arabia : mosquito vector studies. *Med Vet Entomol*, 2002. **16** (4) : 464.
61. Wesson, D. M. et al., Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera : Culicidae). *Mol Phylogenet Evol*, 1992. **1** (4) : 253-69.
62. Brunhes, J. et al., Les moustiques de l'Afrique méditerranéenne, 2000. IRD Montpellier, IRD diffusion.
63. Diallo, M., Ecologie et transmission d'arbovirus à vecteurs culicidiens au Sénégal. Thèse de Doctorat Universitaire, Dakar (Sénégal), Université Cheikh Anta Diop : 2000. 126.
64. Jupp, P. and A. Cornel, Vector competence test with Rift Valley fever virus and five South African species of mosquitoes. *J Am Mosq Control Assoc*, 1988. **4** (1): 4-8.